STATISTICAL RESEARCH REPORT No 1
Institute of Mathematics 1972
University of Oslo

STEPWISE STATISTICAL PROCEDURES WITH GOOD PERFORMANCE IN STEPWISE REDUCED PARAMETER SPACES

by

Erling Sverdrup

1. Introduction.

When constructing a model a statisticians prime concern would be to involve the parameters which interest him (e.g. the overall treatment effect in analysis of variance models). However, in order to make the model realistic, he would have to add "structures" containing additional unknown parameters (e.g. variance of errors, block effects, interactions). The more such parameters he adds, the safer he would feel of having a realistic model. On the other hand these parameters very often are nuisance in the sense that they reduce the usefulness of the model and make inferences about interesting parameters uninteresting and uncertain. Hence it is important to eliminate the "nuisance" parameters - to the extent to which this is justified by the statistical data - and then say something about some of the remaining parameters.

Such problems have been delt with by T.W. Anderson [1], who has developed optimum procedures, and by Robert W. Hogg in [2], [3] and [4]. The present discussion is intended to be somewhat broader in scope.

It is presumed that some of the parameters in the model can be ordered, $\gamma_r, \gamma_{r-1}, \dots$ such that that it is first of all of interest to eliminate γ_r (e.g. because it represents "the highest degree of nuisance"). Only if γ_r is eliminate is it of interest to eliminate γ_{r-1} , etc. The last parameters in the sequence may, in some situations, be those for which the statistical experiment was designed.

Thus we have a situation where the decisions are taken in stages, based on the same set of observations. Such stagewise decisions will often be equivalent to stagewise reduction of the parameter space (e.g. by setting the interactions equal to 0 , etc.).

Some statisticans may feel that, as a matter of principle, the construction of the procedure at the successive stages should be based on the assumptions of reduced parameter spaces. Thus, not only should the distributions of the observations be conditional on the acceptance of the decision at the previous stage, but the class of such distributions should be reduced to a class generated by varying the parameters in the reduced parameter space.

Such a principle would be hard to justify in general. A decision at a certain stage is subject to stochastic uncertainties, and these uncertainties would be disregarded if the reduced class of a priori distributions is taken as a basis at the succeeding stage. The principle would violate the general principle, namely, that any multiple statistical procedure should be jugded from the characteristics of the performance function (i.e. the probabilities of the different combined decisions for the different distributions). Its behaviour ought to be explored for any decision and any distribution of the observations which could not a priori be excluded.

However, there may be situations where it does not matter if the performance is poor for certain combinations of distributions and decisions. In the case of multistage procedures the situation may be such that, having made a mistake at one stage, it does not matter how wrong we are at the succeeding stages. T.W. Anderson [1] gives an example of such a situation. When testing the degree of a polynomial he starts with the term of highest degree and proceeds stagewise. Only if he has correctly jugde the term of degree n to be 0, is he interested in a nice powerfunction for testing the term of degree $n-1$, etc.

We shall be interested in situations where such an attitude is justified. The principle involved could be termed the principle of reduced parameter space as opposed to the principle of retained parameter space. Examples of the last mentioned principle have been given by Lehmann [5].

2. The general problem.

We assume that the observed sample point X has a cumulative distribution function $F_{\tau,\gamma}(x)$ depending on two parameters τ and γ, which a priori can assume any values in sets Θ and Ω respectively. We are interested in $\gamma \in \Omega$, and we assume that

$$\Omega = \omega_{r+1} \supset \omega_r \supset \dots \supset \omega_1 \tag{1}$$

are such that $\Omega - \omega_r$, $\omega_r - \omega_{r-1}$, ... $\omega_2 - \omega_1$, ω_1 represents decreasing degree of complexity of the model. Such a situation would usually arise when $\gamma = (\gamma_1, \dots, \gamma_r)$ and

$$\omega_j = \{\gamma | \gamma_i = 0;\ i=j,\ j+1, \dots r\};\ j=r,\ r-1, \dots, 1 \tag{2}$$

Hence

$$\Omega_j = \omega_{j+1} - \omega_j = \{\gamma | \gamma_i = 0;\ i=j+1, \dots r,\ \gamma_j \neq 0\} \tag{3}$$

and γ_r, $\gamma_{r-1}, \dots \gamma_1$ represents decreasing degrees of complexities in the model. The different γ_j may be vectors with different numbers of components.

It is desired to make a choice between decisions

$$d_r = \text{"rejection of } \gamma \in \omega_r\text{"}; \tag{4}$$

$$d_j = \text{"rejection of } \gamma \in \omega_j\text{, but no rejection of } \gamma \in \omega_{j+1}\text{"},$$
$$j = r-1, \dots, 1 .$$

Thus in the case of a two-way lay out $\gamma_2 = \{\gamma_{ij} | i=1,2,\dots,I;\ j=1,2,\dots,J\}$, $\gamma_1 = \{\alpha_i | i=1,2,\dots,I\}$, $\tau = \{\beta_j;\ j=1,2,\dots,J\}$, where γ_{ij} is the interaction, α_i the treatment effect and β_j the block effect. We want to know (d_2) if there are interaction, and if not, then we want to know (d_1) if there are effects of treatment.

A multiple decision procedure is constructed by defining a partitioning of the sample space in acceptance regions for $d_o, d_1, \ldots, d_r$, respectively. Let $\psi_j(x)$ be the indicator function of the acceptance region for d_j. More generally $\psi_j(x)$ could be the conditional probability of accepting d_j given X,

$$\sum_{j=0}^{r} \psi_j(X) = 1 \tag{5}$$

The unconditional probability of accepting d_j is then $E\psi_j$. Obviously the performance function $E\psi_j$ should be small if $\gamma \in \omega_j$. We shall require

$$E\psi_j(x) = p_j \quad \text{if} \quad \gamma \in \omega_j \text{ ; } j = 1,2,\ldots,r \text{ ;} \tag{6}$$

where $p_1,\ldots,p_r$ are small numbers. It follows that

$$E\psi_o = 1-p_1-\ldots-p_r = p_o \quad \text{if} \quad \gamma \in \omega_1 \text{ ,} \tag{7}$$

since in that case $\gamma \in \omega_j$ for all j.

It is also obvious that one would like

$$E\psi_j \text{ to be large if } \gamma \in \Omega_j = \omega_{j+1}-\omega_j \tag{8}$$

Thus in the case where the ω_j are given by (2), the requirements (6) and (8) amounts to saying that <u>the probability of stating that $\gamma_j \neq 0$ should be large if γ_j is the most complicating parameter; and the same probability should be small if the most complicating parameter present is less complicating than γ_j.</u> This is the fundamental idea behind the principle of <u>reduced parameter</u> space. If a method with <u>retained parameter</u> space is desired, then one would want the probability of stating that $\gamma_j \neq 0$ large in any case; even if $\gamma_i \neq 0$, $i > j$; and one would want the same probability to be small if $\gamma_j = 0$ for any values of the other γ_i, even those for which $i > j$.

Even if the attitude of reduced parameter space might not be quite adequate, it might be adopted anyhow; because it leads to manageable statistical procedures with nice properties.

3. Links with test procedures.

Consider for a moment the problem of testing the hypothesis $\gamma \in \omega_j$ against $\gamma \in \omega_{j+1}-\omega_j$. Let $\delta_j(X)$ be a test function; i. e. an indicator function for the rejection region, or, more generally, the conditional probability of rejecting $\gamma_j = 0$ given X. We may apply the tests δ_j ; $j = 1,2,\dots r$; to construct a decision procedure ψ as follows. First, we use δ_r to decide if d_r should be adopted; i.e. if $\gamma \in \omega_r$ should be rejected. If d_r is not adopted, then we apply δ_{r-1} to decide if d_{r-1} should be adopted, i.e. if $\gamma \in \omega_{r-1}$ should be rejected, etc. Thus

$$\begin{aligned} \psi_r &= \delta_r \\ \psi_{r-1} &= (1-\delta_r)\delta_{r-1} \\ &\text{----------} \\ \psi_j &= \delta_j \prod_{i=j+1}^{r} (1-\delta_i) \\ &\text{----------} \\ \psi_0 &= \prod_{i=1}^{r} (1-\delta_i) \end{aligned} \tag{9}$$

Note that $(j > 0)$

$$1 - \sum_{j}^{r} \psi_i = \prod_{j}^{r}(1-\delta_i) \tag{10}$$

from which we see that (5) holds. Furthermore

$$\psi_j = \delta_j(1-\sum_{j+1}^{r} \psi_i) \tag{11}$$

It is seen from (9) that to any sequence of tests $\delta_1,\dots,\delta_r$, a decision procedure $\psi = (\psi_0,\dots,\psi_r)$ can be constructed. Vice versa, it follows from (11) that to any $\psi = (\psi_0,\dots,\psi_r)$ a sequence of tests $\delta_1,\dots,\delta_r$ can be found, (which may, however, not be uniquely determined from ψ).

4. The fundamental theorem.

We shall say that a multiple procedure $\psi = (\psi_0, \psi_1, \dots, \psi_r)$ has <u>multiple level</u> $(p_1, p_2, \dots, p_r)$ if

$$E\psi_j = p_j \; ; \; \gamma \in \omega_j \; ; \; j = 1,2,\dots,r \tag{12}$$

and that it is <u>performance unbiased</u> with level $(p_1, p_2, \dots, p_r)$ if, in adition,

$$E\psi_j \geq p_j \; ; \; \gamma \in \omega_{j+1} \; ; \; j = 1,2,\dots,r \tag{13}$$

It follows from the discussion above that it is no restriction on the procedure ψ to assume that it has been constructed from tests $\delta_1, \dots, \delta_r$ by means of (9). Furthermore, it is natural to construct the test δ_j for the null hypothesis $\gamma \in \omega_j$ by assuming a priori that $\gamma \in \omega_{j+1}$. We know that in some typical situations in statistics there exists a complete and sufficient statistic D_j under ω_{j+1} and that there is a uniformly most powerful unbiased test of $\gamma \in \omega_j$ which depends on X only through D_j ; $j = 1,2,\dots$.

The following theorem with proof summarizes results and arguments due to <u>T.W. Anderson</u> [1].

<u>Theorem.</u> <u>(i)</u> Assume that $\omega_1 \subset \omega_2 \subset \dots \subset \omega_r \subset \omega_{r+1} = \Omega$ is a sequence of subsets of the space Ω of all values of the component parameter γ. Let D_j be a complete and sufficient statistic for the class of distributions

$$\{F_{\tau,\gamma}\}_{\tau \in \Theta,\, \gamma \in \omega_{j+1}} \tag{14}$$

; $j = 1,2,\dots r$.

<u>(ii)</u> Let $\delta_1, \dots, \delta_r$ be similar tests for the null hypotheses $\omega_1, \dots, \omega_r$, respectively, with levels

$$\alpha_r = p_r \; ; \; \alpha_j = \frac{p_j}{1 - p_{j+1} - \dots - p_r} \; ; \; j = r-1, \dots 2, 1 \; ; \tag{15}$$

respectively. Furthermore, δ_j depends on X only through D_j ; $j = 1,2,\dots r$, and ψ is constructed from $(\delta_1,\dots,\delta_r)$ by means of (9). Then ψ has multiple level $(p_1,p_2,\dots,p_r)$.

<u>(iii)</u> If, in addition, the δ_j are unbiased, then ψ is performance unbiased.

<u>(iv)</u> If, furthermore, δ_j is the uniformly most powerful unbiased test with level α_j ; $j = 1,2,\dots,r$; then the corresponding ψ is such that

$$E\psi_j \geq E\psi_j' \; ; \; \gamma \in \omega_{j+1} \; ; \; j = 1,2,\dots,r \tag{16}$$

where $\psi' = (\psi_0',\psi_1',\dots,\psi_r')$ is any performance unbiased procedure with level $(p_1,\dots,p_r)$. This statement is still true if unbiasedness is left out both as a condition about the δ_j and a condition on ψ' .

<u>Proof of (ii) and (iii).</u> We use proof by induction. Suppose that (12) is true with j replaced by $r,r-1,\dots,j+1$. Then if $\gamma \in \omega_{j+1}$, $E\sum_{j+1}^{r}\psi_i = \sum_{j+1}^{r} p_i$. Since D_j is complete and sufficient under ω_{j+1} , we have $E[\sum_{j+1}^{r}\psi_i|D_j] = \sum_{j+1}^{r} p_r$. Hence by (11) and since δ_j depends only on D_j ,

$$E\psi_j = E\delta_j E[(1-\sum_{j+1}^{r}\psi_i)|D_j] = (1-\sum_{j+1}^{r}p_i)E\delta_j \tag{17}$$

provided $\gamma \in \omega_{j+1}$. If $\gamma \in \omega_j$ then $E\delta_j = \alpha_j$ and (12) follows from (17) and (15). If $\gamma \in \omega_{j+1}$ then $E\delta_j \geq \alpha_j$ and (13) follows in the same manner. Obviously (12) and (13) respectively are true with $j = r$, since $\psi_r = \delta_r$ and $p_r = \alpha_r$. Hence (12) and (13) respectively follow by induction.

<u>Proof of (iv).</u> Let ψ' be an arbitrary performance unbiased procedure with level $(p_1,\dots,p_r)$ and let ψ be a procedure satisfying the assumptions in (ii) and (iii) (but not necessarily the assumption

in (iv)). Furthermore, let δ_h^* be a uniformly most powerful unbiased test for the nullhypothesis ω_j against $\omega_{j+1}-\omega_j$, which depends on X only through D_h. We define

$$\psi_h^* = \delta_h^*(1-\sum_{h+1}^{r}\psi_i) \tag{18}$$

and shall prove that $E\psi_h^* \geq E\psi_h'$ if $\gamma \in \omega_{h+1}$. (iv) is an immediate consequence of this result; which is, in fact, a little more general.

We note from (17) and (11) that

$$E\delta_h^* \sum_{h+1}^{r} \psi_i = (\sum_{h+1}^{r} p_i)E\delta_h^* \tag{19}$$

if $\gamma \in \omega_{h+1}$. Hence, since ψ and ψ' are perfermance unbiased and δ_h^* is power unbiased,

$$E(\psi_h'+\delta_h^*\sum_{h+1}^{r}\psi_i) = E\psi_h'+\sum_{h+1}^{r}p_iE\delta_h^* \geq p_h+\sum_{h+1}^{r}p_i\alpha_h = \alpha_h \tag{20}$$

with equality if $\gamma \in \omega_h$. Thus $\psi_h'+\delta_h^*\sum_{h+1}^{r}\psi_i$ could be considered as an unbiased level α_h test for ω_h; but since δ_h^* is the uniformly most powerful among those tests, we have

$$E\delta_h^* \geq E(\psi_h'+\delta_h^*\sum_{h+1}^{r}\psi_i) , \tag{21}$$

which, by (18), could be written $E\psi_h^* \geq E\psi_h$. The last statement in (iv) is proved by going through the proof once more, leaving out the inequality in (20) and any reference to unbiasedness. Q.E.D.

We remarked above that the result we have just proved is a little more general than (iv). It states that even if we have used a nonoptimum test at the previous r-h stages, we should still use the ordinary optimum test at the next stage.

5. Notes to the theorem.

Note A. The level. About the choice of $p_1,\dots,p_r$ the following should be noted. Since the statements "$\gamma \notin \omega_j$" are the only statements that can be false, the probability of making a false statement when $\gamma \in \omega_j$ is $p_j+\dots+p_r$ and the maximum probability of a false statement is $p_1+\dots+p_r = 1-p_o$. For convenience we might choose $\alpha_i = \alpha$ indepently of i . Then $p_j = \alpha(1-\alpha)^{r-i}$, $j > 0$, and $p_o = (1-\alpha)^r$. It would perhaps be reasonable to choose $1-p_o = \epsilon(=0.05 \text{ or } 0.01)$ and then $\alpha = 1-\sqrt[r]{1-\epsilon}$.

Note B. Independent component tests. It is easily seen that the result in (ii) and (iii) of the theorem are true if any reference to the complete and sufficient statistics D_j ; $j = 1,2,\dots r$; is left out in the assumption and it is assumed instead that $\delta_r,\dots,\delta_j$ are stochastically independent when $\gamma \in \omega_{j+1}$; $j = r-1,\dots,1$.

There is, in fact, a close connection between this assumption of independence and the assumption involving sufficiency and completeness. This is seen as follows. The requirement that δ_g is similar, i.e. $E\delta_g = \alpha_g$ everywhere in ω_g , is usually brought about by δ_g having a distribution which is independent of $\gamma \in \omega_g$ (e.g. the Student statistic is independent of the variance). Hence for $j < g$ the distribution of δ_g is independent of $\gamma \in \omega_{j+1}$. On the other hand D_j is sufficient and complete under ω_{j+1} . Then it is known from a theorem of Basu that D_j and δ_g are independent. Since δ_j depend on X only through D_j , it follows that also δ_g and δ_j are independent if $\gamma \in \omega_{j+1}$. By stepwise application of this result, starting with $g = r$, we find that $\delta_r,\delta_{r-1},\dots,\delta_j$ are independent if $\gamma \in \omega_{j+1}$.

Note C. The assumptions of the theorem are fulfilled in the particular case of a regular Darmois-Koopman exponential class of distributions. The probability element of X is given by

$$dF_{\tau,\gamma}(x) = A(\tau,\gamma)e^{\sum_1^s \tau_i T_i(x) + \sum_1^r \gamma_j C_j(x)} dF_0 . \tag{22}$$

$0 \in \Theta$ and $0 \in \Omega$ (where Θ and Ω are define above) are assumed to be inner points in Θ and Ω respectively. F_0 is the distribution of X when $\tau = 0$, $\gamma = 0$. If $\omega_1,\dots,\omega_r$ are defined by (2), then the existence of a uniformly most powerful test δ_j for $\gamma \in \omega_j$ against $\gamma \in \omega_{j+1}-\omega_j$, is guaranteed.As a matter of fact, with $D_j = (T, C_1,\dots,C_j)$, the δ_j is given by

$$\begin{aligned} \delta_j &= 1 \quad \text{if} \quad C_j > g_{2j}(D_{j-1}) \quad \text{or} < g_{1j}(D_{j-1}) \\ \delta_j &= 0 \quad \text{if} \quad g_{1j}(D_{j-1}) < C_j < g_{2j}(D_{j-1}) \end{aligned} \tag{23}$$

where g_{1j} , g_{2j} and $\delta_j(X)$ when $C_j = g_{ij}(D_{j-1})$ are determined such that

$$E(\delta_j | D_{j-1}) = \alpha_j = \frac{p_j}{1-p_{j+1}-\dots-p_r} \tag{24}$$

when $\tau = 0$, $\gamma = 0$, i.e. when the distribution of X is F_0 .

Note D. One sided alternatives. Sometimes one would want to test $\gamma_1 = \gamma_2=\dots=\gamma_r = 0$ against one-sided alternatives, i.e. it is a priori assumed that $\gamma_j \geq 0$. The Ω_j in (3) should then be replaced by

$$\Omega_j = \{\gamma | \gamma_{j+1}=\dots=\gamma_r = 0 \ , \ \gamma_j > 0\} \tag{25}$$

In the case of the Darmois-Koopman family of distributions (22), the δ_j given by (23) should be replaced by a δ_j defined by

$$\begin{aligned} \delta_j &= 1 \quad \text{if} \quad C_j > g_j(D_{j-1}) \\ \delta_j &= 0 \quad \text{if} \quad C_j < g_j(D_{j-1}) . \end{aligned} \tag{26}$$

Furthermoe, there would be no difficulties in operating with "mixed" situations, some γ_j being tested one-sided and some two-sided.

6. Testing of fixed mean effects in the normal case.

Suppose that the components of $X = (X_1, \dots, X_n)'$ are independent and normally distributed with unknown variance $\sigma^2 = \text{var } X_j$ and means $\xi = (\xi_1, \dots, \xi_n)' = y\beta$, where $y = \{y_{ij}\}$ is a matrix and $\beta = (\beta_1, \dots, \beta_q)'$ a priori is subject to a set of linear restrictions represented by $g\beta = 0$. Let these a priori assumptions amount to restricting ξ to vary freely in the $n - m$ dimensional space, i.e. they represent m independent linear restrictions on ξ . Furthermore, let $g_1\beta = 0$, $g_2\beta = 0, \dots g_r\beta = 0$ represent respectively $m_1, m_2, \dots, m_r$ further restrictions on ξ , such that the $m_1 + \dots + m_{r+1}$ restrictions are independent $(m_{r+1} = m)$.

First of all we want to test H_r that $g_r\beta = 0$. If H_r is not rejected we want to test H_{r-1} that $g_{r-1}\beta = 0$ etc. We shall adopt the attitude inherent in the method of reduced parameter space. In principle this would mean that if some of the hypotheses $H_r, \dots, H_{j+1}$ are false, then we do not care so much if we fail to reject H_j if H_j is wrong or mistakenly reject H_j if H_j is correct. In case of two stages, $r = 2$, we shall show, however, that if we use the ψ defined by (23) and (24), then we have complete control of the multiple level (p_1, p_2) even from the point of view of a retained parameter space, i.e. $E\psi_1 \leq p_1$ for $g_1\beta = 0$ also when $g_2\beta \neq 0$.

The above set-up should cover most situations of multistage testing of the means in the case of fixed effects connected with the normal distributions.

Suppose, now, that Q_j is the minimum of $Q = (X-\xi)'(X-\xi)$ with respect to ξ under the restrictions $g_k\beta = 0$; $k = j, \dots, r+1$,

$(g_{r+1} = g)$. Let δ_j be 1 if

$$(Q_j - Q_{j+1}) \sum_{j+1}^{r+1} m_k / m_j Q_{j+1} > f_j \tag{27}$$

and 0 otherwise, where f_j is the $1 - \alpha_j$ fractile of the Fisher distribution with m_j and $\sum_{j+1}^{r+1} m_k$ degrees of freedom. Then it is known that regarding δ_j as a test of H_j under the "a priori" assumptions $g_k\beta = 0$; $k = j+1,\ldots,r+1$; it is unbiased with level α_j and depends on X only through a set of statistics which is sufficient and complete when $g_k\beta = 0$; $k = j+1,\ldots r+1$. From item (i)-(iii) of the theorem it then follows that the ψ corresponding to $(\delta_1,\ldots,\delta_r)$ see eq. (9)) is performance unbiased with multiple level $(p_1,\ldots,p_r)$, if

$$\alpha_j = p_j/(1-\sum_{j+1}^{r} p_k) \; ; \; j = 1,2,\ldots r \, .$$

In the case when $m_1 = m_2 = \ldots = m_2 = 1$, all the δ_j are Student's tests (with $m+r-j$ degrees of freedom) and they are known to be uniformly most powerful if $g_k\beta = 0$; $k = j+1,\ldots r+1$. It follows from the theorem item (iv) that the corresponding ψ uniformly maximizes the performance, in the restricted spaces, among all performance unbiased procedures.

Expanding somewhat on the assertions above, it is well known that the situation described above can be given the following canonical form by a transformation of X to variables Y_{vj} ; $v = 1,2,\ldots m_j$; $j = 1,2,\ldots,r+1$; Z_v ; $v = 1,2,\ldots,N$; where the $n = m_1+\ldots+m_{r+1}+N$ variables are independent and normal with variance σ^2 and

$$EY_{vj} = \gamma_{vj} \; ; \; j = 1,2,\ldots r \; ; \; EY_{vr+1} = 0 \; , \; EZ_v = \tau_v \tag{28}$$

for all v. Let $Y_j = (Y_{1j},\ldots,Y_{m_j j})'$, $\gamma_j = (\gamma_{1j},\ldots,\gamma_{m_j j})' = EY_j$

and $\tau = (\tau_1,\ldots,\tau_N)'$. Then $g_k\beta = 0$; $k = j,\ldots,r+1$ is equivalent to $\gamma_k = 0$; $k = j,\ldots,r$, i.e. $\gamma \in \omega_j$ where ω_j is defined by (2) and $\gamma = (\gamma_1',\ldots,\gamma_r')'$. From the joint density of all Y_{vj} and Z_w it follows that

$$Q_j = \sum_j^{r+1} Y_j'Y_j \text{ , hence } Q_j-Q_{j+1} = Y_j'Y_j \text{ .} \tag{29}$$

On the other hand we can write the probability element in the form (22), with the γ_j now vectors, $(C_1,\ldots,C_r) = (Y_1',\ldots,Y_r')$ and

$$(T_1,\ldots,T_N) = (\sum_1^{r+1} Y_j'Y_j+\sum_1^N Z_v^2, Z_1,\ldots,Z_N)$$

If $\gamma \in \omega_{j+1}$, i.e. $\gamma = 0$; $k = j+1,\ldots r$; then $(T_1,\ldots,T_N,C_1,\ldots C_j)$ is complete and sufficient. It is one-to-one corresponding with

$$D_j = (Q_j,Y_1',\ldots,Y_j',Z_1,\ldots,Z_N) \tag{30}$$

Using this D_j in the theorem ,(i)-(iii), we obtain the assertion that ψ defined by (27) is performance unbiased with multiple level $(p_1,\ldots,p_r)$. If $m_1 = m_2=\ldots=m_r = 1$ then it follows from the remarks made in section 5, note C, that the δ_j defined by (27), which can be written

$$Y_j^2(m+r-j)/Q_{j+1} > f_j$$

is the uniformly most powerful unbiased test and hence, by the theorem item (iv), the corresponding ψ maximizes the performance uniformly, as asserted above.

We shall now show that if $r = 2$, <u>then</u> $E\psi_1 \leq p_1$ <u>any β such that</u> $g_1\beta = 0$. In fact, we <u>shall show that</u> $E\psi_1$ <u>is a decreasing</u> function of

$$\lambda_2 = \gamma_2'\gamma_2/\sigma^2 = \frac{1}{\sigma^2} Q_2 |_{x=\xi} \tag{31}$$

and hence is maximum for $\gamma_2 = 0$; i.e. $g_2\beta = 0$, in which case it is equal to p_1 . Let $q_j = Q_j - Q_{j+1}$ and $k_j = f_j m_j / \sum_{j+1}^{r+1} m_k$. Then

$$E\psi_1 = \Pr(q_2 \leq k_2 q_3 \cap q_1 > k_1(q_2+q_3)) =$$

$$= \Pr(q_2 \leq \min(k_2 q_3 , \frac{q_1}{k_1} - q_3)) \tag{32}$$

Let $H(w) = \Pr(W \leq w)$, where

$$W = \frac{1}{\sigma^2} \min(k_2 q_3, \frac{q_1}{k_1} - q_3) \tag{33}$$

q_3/σ^2 and q_1/σ^2 are independent with distributions which are independent of all parameters. Hence $H(w)$ does not depend on β and σ . Since, furthermore, q_2/σ^2 is independent of q_3 , q_1 , we get

$$E\psi_1 = \int_0^\infty \Gamma_{m_2}(w ; \lambda_2) dH(w) \tag{34}$$

where Γ_{m_2} is the eccentric chi-square distribution with m_2 degrees of freedom and eccentricity λ_2 . However, this intergrand is known to be decreasing function of λ_2 . The assertion follows.

Example 1. Testing interaction and treatment effects. Let X_{ijv} ; $v = 1,2,\ldots m$; $i = 1,2,\ldots,I$; $j = 1,2,\ldots,J$; be independent normal with variance σ^2 and expectation $\xi + \alpha_i + \beta_j + \gamma_{ij}$ $(\Sigma\alpha_i = \Sigma\beta_j = \sum_j \gamma_{ij} = \sum_i \gamma_{ij} = 0)$. We want to know if we can state (d_2) that some $\gamma_{ij} \neq 0$. In case we cannot, we want to know if we can state (d_1) that some $\alpha_i \neq 0$. The method of reduced parameter space is to the effect of being interested in a good test of $\gamma = \{\gamma_{ij}\} = 0$ against $\gamma \neq 0$ under all circumstances, but in a good test of $\alpha = \{\alpha_i\} = 0$ against $\alpha \neq 0$ only if $\gamma = 0$. Thus if there are interaction we are not interested in the main effect of treatment and if mistakenly we should test the main effect in the case where there is interaction, then we do not care if the test is poor. In order to avoid misunderstanding it should also be stressed that we do not test interaction for the purpose, only, of neglecting it when testing treatment effect. There must be an interest in interaction for its own sake.

Let now

$$\begin{aligned}
&\overline{X}_{ij} = \frac{1}{m}\sum_{v}\overline{X}_{ijv}\ ,\ \overline{X}_{i\bullet} = \frac{1}{J}\sum_{j}\overline{X}_{ij}\ ,\ \overline{X}_{\bullet j} = \frac{1}{I}\sum_{i}X_{ij}\ ,\\
&\overline{X} = \frac{1}{n}\sum_{i,j,v}X_{ijv}\\
&Q_o = \sum_{i,j,v}(X_{ijv} - \overline{X}_{ij})^2\ ,\ Q_{AB} = m\sum_{i,j}(\overline{X}_{ij}-\overline{X}_{\bullet i}-\overline{X}_{j\bullet}+\overline{X})^2\\
&Q_A = mJ\sum_{i}(\overline{X}_{i\bullet} - \overline{X})^2\ ,\ \nu = IJ(m-1)\ ,\\
&\mu = (I-1)(J-1)\ ,\ n = IJm
\end{aligned} \tag{35}$$

The multple procedure ψ defined by (27) leads to stating that some $\gamma_{ij} \neq 0$ if

$$F_{AB} = Q_{AB}\nu/Q_o\mu > f_2$$

where f_2 is the $1-p_2$ fractile of the Fisher distribution with μ and ν degrees of freedom. Thus $E\psi_2 = p_2$ if all $\gamma_{ij} = 0$.

Furthermore, ψ leads to stating that some $\alpha_i \neq 0$ if $F_{AB} \leq f_2$ and

$$F_A = Q_A(\nu+\mu)/(Q_o+Q_{AB})(I-1) > f_1 \tag{36}$$

where f_1 is the $1-\alpha_1'$ fractile of the Fisher distribution with $I-1$ and $\nu+\mu$ degrees of freedom and $\alpha_1' = p_1/(1-p_2)$. Thus $E\psi_1 = p_1$ if all $\gamma_{ij} = 0$ and all $\alpha_i = 0$. Referring to section 5 note B this determination of the levels also follows from the fact that, by Basu's theorem, F_{AB} and F_A are independent if all $\gamma_{ij} = 0$.

This method is, by the theorem, performance unbiased with multiple level (p_1, p_2) . If $I=J=2$ it has, moreover, uniformly maximum performance function in the reduced parameter spaces. [The complete and sufficient statistics referred to in the theorem are in this special case respectively $D_2 = (\overline{X}_{11},\dots,X_{IJ},Q_o)$ and $D_1 = (\overline{X}_{1\cdot},\dots \overline{X}_{I\cdot}$, $\overline{X}_{\cdot 1},\dots,X_{\cdot J}$, $Q_o+Q_{AB})]$.

Let us consider the case $I = J = 2$. There are then three scalar parameters $\alpha = \alpha_1 = -\alpha_2$, $\beta = \beta_1 = -\beta_2$, $\gamma = \gamma_{11} = \gamma_{21} = \gamma_{12} = -\gamma_{21}$ in the means. We want to test if $\gamma \neq 0$ (i.e. test $\gamma = 0$), and in case it is not,we want to know if $\alpha > 0$, (i.e. test $\alpha \leq 0$) . Let S^2 be the usual unbiased estimate of σ^2, i.e. $(n-4)S^2 = Q_o$, and $\hat{\alpha}$ and $\hat{\gamma}$ the least square estimates of α and γ . Then by the principle of reduced parameter space we should state (d_2) that $\gamma \neq 0$ if

$$n\hat{\gamma}^2 > S^2t^2 \tag{37}$$

where t is the $1 - \frac{p_2}{2}$ fractile of the Student distribution with n-4 degrees of freedom. If this inequality is not fulfilled then we should state that $\alpha > 0$ in case

$$\sqrt{n}\,\hat{\alpha} > S't' \tag{38}$$

where t' is the $1-\alpha_1$ fractile $(\alpha_1 = p_1/(1-p_2))$ of the Student distribution with $n-3$ degrees of freedom and

$$(n-3)S'^2 = (n-4)S^2+n\hat{\gamma}^2 \tag{39}$$

This method will, by what we have proved above, have multiple level (p_1, p_2) also if we retain the parameter space at the second level.

No efforts have, however, been made to attain high sensitivity against alternatives $\alpha > 0$, $\gamma \neq 0$ (only against $\alpha > 0$, $\gamma = 0$), such as with the method suggested by Lehmann [5]. Lehmann's method can be given the following form. Use the same method as above to test if $\gamma = 0$ should be rejected. In case not, then state that $\alpha > 0$ if $\sqrt{n}\,\hat{\alpha} > St(W)$, where $t(W)$ is the $1-\alpha_1(W)$ fractile of the Student distribution with $n-4$ degrees of freedom,

$$\alpha_1(W) = \alpha_1+(1-2\alpha_1)G_{n-4}(\sqrt{(n-4)(t^2W-1)})$$

$$W = [(n-4)S^2+n\hat{\alpha}^2]/n\hat{\gamma}^2 \tag{40}$$

Thus the method, which is really conditionally on $n\hat{\gamma}^2 < S^2t^2$, could easily be performed "unconditionally" with an "unconditional stochastic level" $\alpha_1(W)$. The conditional level is, of course, α_1. Note that by this method we use $n-4$ degrees of freedom also at the second stage, not $n-3$ degrees of freedom as with the method of reduced parameter space.

7. Some more examples.

Example 2: Testing heteroscedasticity and treatment effect.

X_{jv} ; $v = 1,2,\ldots,n_j$; $j = 1,2,\ldots s$ are independent normal, var $X_{jv} = \sigma_j^2$, $EX_{jv} = \xi_{jv}$. We want to find out if $\sigma_1^2,\ldots,\sigma_s^2$ are different and in case we cannot make such a statement we sould like to find out if $\xi_1,\ldots \xi_s$ are different. Again it should be emphasized that we are not interested in testing $\sigma_1=\ldots=\sigma_s$ for the purpose of providing justification for using ordinary analysis of variance in testing $\xi_1=\ldots=\xi_s$. Our attitude is that we are only interested in the hypothesis $\xi_1=\ldots=\xi_r$ if the σ_j^2 are equal, even to the extent of requiring no good properties of the test if mistakenly we should test it when the σ_j^2 are different.

Applying the theorem of section 4, it is seen that the complete and sufficient set of statistics a priori D_2 is

$$\overline{X}_j = \frac{1}{n_j}\sum_v X_{jv} \; ; \; n_jS_j^2 = \sum_v (X_{jv}-\overline{X}_j)^2 \; ; \; j = 1,\ldots s \; .$$

If $\sigma_1=\ldots=\sigma_s$ the complete and sufficient set of statistics is

$$\overline{X}_j \; ; \; j = 1,2,\ldots,s \; ; \; nS^2 = \sum_{j,v} (\overline{X}_{jv}-\overline{X})^2$$

where $n = \sum n_j$.

At the first stage we should, obviously, use a test depending on $S_1^2,\ldots S_s^2$; e.g. the likelihood ratio test with critical region

$$V_2 = S^{2n}/\prod_1^s S_j^{2n_j} > v_2 \tag{41}$$

Since V_2 has a distribution which is independent of the common variance and $\xi_1,\ldots,\xi_s$, we can adjust v_2 to a level v_2 . Numerically Bartlett's approximations are used. Alternatively, in the case when $n_1=\ldots=n_s$, we may use

$$V_2' = \max S_j^2/\min S_j^2 > v_2' \,, \tag{42}$$

which may easily be adjusted to the level α_2 .

At the second stage the ordinary analysis of variance is used, with critical region

$$V_1 = (n-s)\sum_1^s n_j(\overline{X}_j-X)^2/n(s-1)S^2 > v_1 \tag{43}$$

where $\overline{X} = \frac{1}{n}\sum_{j,v} X_{jv}$ and v_1 is the $1-\alpha_1$ fractile of the Fisher distribution with $s-1$ and $n-s$ degrees of freedom.

The method has, by the theorem item (i)-(ii), multiple level $(p_1,\ p_2)$; and - at least when critical region $V_2' > v_2'$ is used at the first stage - it has unbiased performance by item (iii) of the theorem.

It should be noted that it follows directly from Basu's theorem that $(\overline{X}_1,\ldots X_s,S^2)$ is independent of $V_2,\ V_2'$.

Example 3. Testing for complexity of treatment in Bernoulli trials in order to escalate quality.

We have $s+1$ Bernoulli trial sequences, $d_0,d_1,\ldots,d_s$, representing respectively controle d_0 and more and more complex treatments $d_1,d_2,\ldots,d_s$ of $n_0,n_1,\ldots,n_s$ units respectively. The probability of a desirable property A of anyone unit treated by d_i is π_i . Treatment d_1 could be thought of as a simple treatment, d_2 as treatment d_1 with something added, treatment d_3 is treatment d_2 with something added etc. It is obvious a priori that $\pi_0 \leqq \pi_1 \leqq \ldots \leqq \pi_s$. We want to know how complex a treatment we have to choose, taking into account that on the one hand side we want to escalate quality measured by π_i and on the other hand side we want as simple treatment as possible. This would mean that we want to find the smallest i (i.e. the least complex d_i) for which $\pi_i = \pi_{i+1} = \ldots = \pi_s$.

However, there may be a certain increase from π_j to π_{j+1} which could be considered too trifling to justify the refinement from d_i to d_{i+1} . Let it be such that the relative increase in odds should be at least $R \geqq 1$ (at any stage), i.e.

$$\rho_i = \frac{\pi_i}{1-\pi_i} \geqq R \frac{\pi_{i-1}}{1-\pi_{i-1}} = R\rho_{i-1} \tag{44}$$

Thus we want to choose between decisions d_i ; $i = 1,2,\ldots,s$; where d_i states that

$$\rho_i > R\rho_{i-1} \text{ , } \rho_{i+1} \leqq R\rho_i \text{ , } \rho_{i+2} \leqq R\rho_{i+1}, \ldots, \rho_s \leqq R\rho_{s-1} \tag{45}$$

Assume that among the n_i items treated with d_i there are Y_i attaining the property A .

	d_o	d_1	d_2	......	d_s
Pr(A)	π_o	π_1	π_2	......	π_s
Number of trials	n_o	n_1	n_2	......	n_s
Number with A	Y_o	Y_1	Y_2	......	Y_s

The probability of a specified outcome is

$$\prod_{i=0}^{s} \binom{n_i}{Y_i} \pi_i^{Y_i} (1-\pi_i)^{n_i - Y_i} = A\Pi\binom{n_i}{Y_i}\rho_i^{Y_i}$$

Let F_o be the probability distribution corresponding to $\rho_o = 1$, $\rho_i = R\rho_{i-1}$; i.e. $\rho_i = R^i$; then it is seen that the probability element of $(Y_o,\ldots,Y_s)$ could written

$$A_o \theta^{\sum_{j=o}^{s} C_j \gamma_j} dF_o \tag{46}$$

where

$$\gamma_j = \log \rho_j - \log \rho_{j-1} - \log R \tag{47}$$

$$C_j = \sum_j^s Y_i \tag{48}$$

Now it is seen that (45) could be written

$$\gamma_i > 0 \;,\; \gamma_j \leqq 0 \;;\; j = i+1,\ldots,s \tag{49}$$

If $\gamma_{i+1}=\ldots=\gamma_s = 0$, then we see from (46) and section 5 note D that we should state with probability 1 that $\gamma_i > 0$ if $C_i > g_i(C_o,\ldots,C_{i-1})$, and with probability $A_i(C_o,\ldots,C_{i-1})$ that $\gamma_i > 0$ if $C_i = g_i(C_o,\ldots,C_{i-1})$, where g_i and A_i are determined such that the conditional probability of stating $\gamma_i > 0$ under F_o given $C_o,\ldots,C_{i-1}$ is α_i. But there is a one-to-one correspondance between $(C_{i-1},\ldots,C_o)$ and $(C_{i-1},Y_{i-2},\ldots,Y_o)$ and (C_i, C_{i-1}) is independent of $Y_{i-2},\ldots,Y_o$. Hence we can write $g_i(C_{i-1})$ and $A_i(C_{i-1})$ in the statements above and we could condition with respect to C_{i-1} only. But $C_i > g_i(C_{i-1})$ could be written $Y_{i-1} < C_{i-1} - g(C_{i-1}) = a_i(C_{i-1})$.

In order to determine a_i and A_i we need the conditional probability of Y_{i-1} given C_{i-1} when $\rho_i = R^i$. We find that

$$q_i(y|c) = Pr(Y_{i-1} = y|C_{i-1} = c) =$$

$$= \binom{n_{i-1}}{y} R^{(i-1)y} Q_i(c-y)/Q_{i-1}(c) \tag{50}$$

where

$$Q_i(c) = \sum_{y_i+\ldots+y_i=c} \binom{n_i}{y_i}\cdots\binom{n_o}{y_o} R^{\sum_i^s jy_j} \tag{51}$$

which can be found from the recursion formulae

$$Q_{i-1}(c) = \sum_{y=o}^{c} \binom{n_{i-1}}{y} R^{(i-1)y} Q_i(c-y) \tag{52}$$

$Q_i(c)$ reduces to $\binom{n_i+\dots+n_s}{c}$ if $R = 1$.

Thus $a_i(c)$ and $A_i(c)$ should be determined by

$$\sum_{y=o}^{a_i(c)-1} q_i(y|c) + A_i(c) q_i(a_i(c)|c) = \alpha_i \tag{53}$$

<u>Having found</u> $a_i(c)$; $A_i(c)$; $i = 1,2,\dots,s$ <u>by means of (50) - (53) our procedure is determined. Having not obtained any statements</u> $\rho_j > R\rho_{j-1}$ <u>for</u> $j = s,s-1,\dots,i+1$, <u>we should state</u> $\rho_i > R\rho_{i-1}$ <u>with probability 1 and</u> $A_i(C_{i-1})$ <u>according as</u> $Y_{i-1} < a_i(C_{i-1})$ <u>or</u> $= a_i(C_{i-1})$. As a "practical" working rule one could start with $i = s$, then $i = s-1$ until we reach the first i (if any) for which

$$\sum_{o}^{y_{i-1}} \binom{n}{y} R^{(i-1)y} Q_i(C_{i-1}-y) < \alpha_i Q_{i-1}(C_{i-1}) \tag{54}$$

The procedure has the following properties.

<u>(i)</u> Pr (escalation to $d_i | \rho_i \leqq R\rho_{i-1}$; $\rho_j = R\rho_{j-1}$; $j = i+1,\dots,s$) $\leqq p_i$

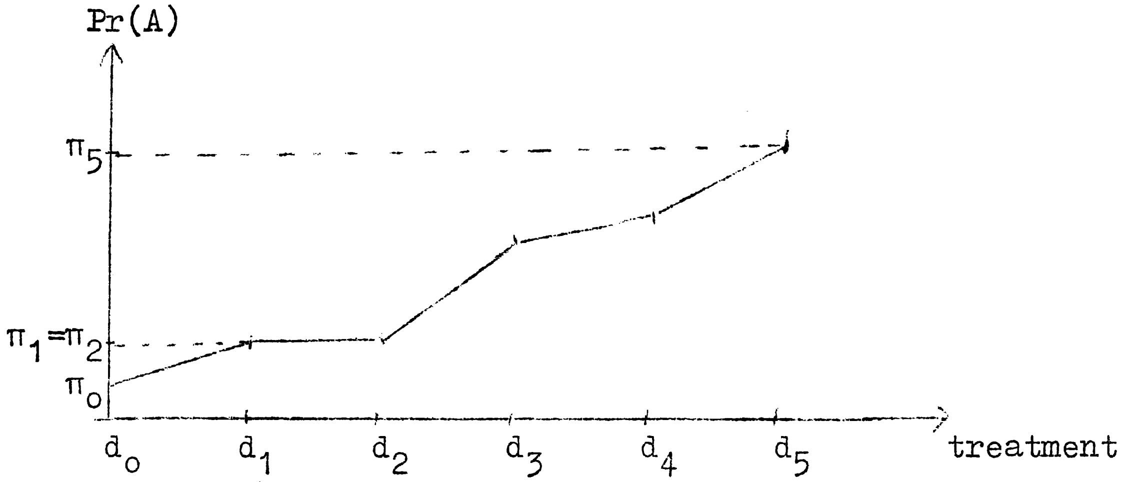
<u>(ii)</u> Pr (escalation to $d_i | \rho_i > R\rho_{i-1}$; $\rho_j = R\rho_{j-1}$; $j = i+1,\dots,s$) $\geqq p_i$

<u>(iii)</u> The probability in (ii) is greater than for any other procedure satisfying (i) and (ii).

Apart from the general objection to the mildness of these requirements connected with the reduced parameter space, there is the special objection in this case that (ii) - (iii) should really be required (at least) for $\rho_j \leqq R\rho_{j-1}$, $j = i+1,\dots,s$. However, considering that the performance function are continuous, that the interval $(1,R^s)$

usually will be short, making the Bernoulli probabilities π_i vary in short intervals $(\frac{\rho_o}{1+\rho_o}, \frac{\rho_o R^i}{1+\rho_o R^i})$; $i = 1,2,\ldots,s$; this objection may not be serious.

The general objection that (i) should really be true for any ρ's such that $\rho_i \leqq R\rho_{i-1}$ and (ii), (iii) should be required whenever $\rho_i > R\rho_{i-1}$; could be illustrated in the following manner, assuming for convenience $R = 1$. Suppose we have 5 treatments with the following true escalation effect.



Statistican I chooses d_1 whereas statistican II chooses d_2 . They would both be wrong, since they have failed to see the escalation from d_2 to d_5 . However, II would be worse off, because he has recommended a more complicated treatment without obtaining any better result. The objection to the requirement (i) is that it does not recognize the destinction between decisions d_1 and d_2 in a situation like the one just described.

It is seen, howver, that if it could be assumed a priori that the quality would have a deminishing (concave) increase, then such a situation could not arise. This would dispose of the objection to the mildness of requirement (i). However, if there were a sharp increase from d_1 to d_2 , then this ought to be discovered, even if it is not the "last" increase. Hence a stronger requirement about sensitivity than (ii) and (iii) would still be wanting.

The method is justified anyhow, since methods which are sensitive to any alternative in the original parameter space seem to be prohibitive from a practical point of view.

REFERENCES

[1] T.W. Anderson: "The choice of degree of a polynomial regression as a multiple decision problem." Ann. Math. Stat., vol. 33 (1962), p. 255.

[2] Robert W. Hogg: "On the resolution of statistical hypotheses." Am. Stat. Ass. Journal 56 (1961), p. 978.

[3] Robert W. Hogg: "Iterated tests of the equality of several distributions." Am. Stat. Ass. Journal 57 (1962), p. 579.

[4] Robert W. Hogg and Elliot A. Tanis: "An iterated procedure for testing the equality of several exponential distributions." Am. Stat. Ass. Journal 58 (1963), p. 435.

[5] E.L. Lehmann: "A theory of multiple decision problems, II." Ann. Math. Stat., vol. 28 (1957), p. 547.